



TITLE:

<研究論文(原著論文)>毒を飲むことは合理的か

AUTHOR(S):

佐藤, 広大

---

CITATION:

佐藤, 広大. <研究論文(原著論文)>毒を飲むことは合理的か.  
Contemporary and Applied Philosophy 2019, 10: 51-65

ISSUE DATE:

2019-04-19

URL:

<https://doi.org/10.14989/240896>

RIGHT:

# 毒を飲むことは合理的か\*

佐藤 広大

## 概要

Gregory Kavka's toxin puzzle is a famous thought experiment. In my view, this puzzle consists of several questions. Among them, I deal with the one question: is it rational to drink the toxin? I reply that, in the toxin puzzle, it is rational to form a belief that drinking the toxin is rational. This reply is based on the idea that an agent who thinks moderately is rational. This idea is in contrast with the traditional dogma that the more an agent thinks, the more rational the agent becomes. In order to insist so, I explain the content of the toxin puzzle (section 2), and introduce arguments against and for insisting that it is rational to drink the toxin (section 3). My argument for insisting that they are rational agents who form a belief that drinking the toxin is rational is based on Richard Holton's argument. He insists that it is a rational disposition not to rethink the intention to drink the toxin in the toxin puzzle. Instead, I insist that it is rational to form the belief that drinking the toxin is rational in the toxin puzzle. At least, I will suggest that the toxin puzzle is worth consideration.

Keywords: 毒パズル, 合理性, 意図, 信念.

## 1 はじめに

哲学の目的はパズルを解くことそれ自体ではない。しかし、パズルを解くことは、哲学的な主題を考えるよい機会になる。パズルは、哲学的な主題を考えるための単純化された実験場としての役割を果たすだけでなく、哲学的な主題が交わるハブとしての役割も果たす。あるパズルの名のもとに、そのパズルと関連する様々な主題が検討され、それと同時に、それら主題同士の関係も明らかになる。

それでは、どのようなことがパズルの豊かさに関する指標となるだろうか。ある一つの主題の解明にどれくらい寄与したかはパズルの豊かさに関する一つの指標となるだろう。そして、どれくらい多くの主題同士

---

\* CAP Vol. 10 (2018-2019) pp. 51-65. 受理日: 2018.07.19 採用日: 2019.02.26 採用カテゴリ: 研究論文(原著論文)  
掲載日: 2019.04.19.

を結びつけたのかということもパズルの豊かさに関する指標となるはずである。

本稿の題材もパズルである。そのパズルとは、Kavka (1983) で提出された「毒パズル (the toxin puzzle)」である。国内ではこれまでほとんど取り上げられることのなかった<sup>\*1</sup>毒パズルだが、意図とは何か、合理性とは何か、哲学と脳科学の知見の間にはどのような連関があるかといった主題それぞれの解明に寄与するだけでなく、そうした主題同士の関係の解明にも寄与する。

本稿は、そうした主題のなかでも、毒パズルにおいて毒を飲むことは合理的かという主題を中心に論じられてきた問題を取り上げる。以下では、さしあたり「合理的」という語を次のような意味で用いる。ある選択肢が合理的なのは、その選択肢をとることが行為者にとって最善で、十分な理由がある場合である。ただし、本稿の後半では、「合理的」という語を選択肢に関してだけでなく、行為者に関しても派生的に用いる。ある行為者が合理的なのは、その行為者にとって自分の状態が最善な場合である。ただし、ある状態が行為者にとって最善だが、その状態が最善である理由を答えることができない場合であっても、その状態は合理的と拡張的に形容されうるということを私は最終的に提案する。

本稿の目標は、毒を飲むことが合理的だという信念を形成することは合理的だと主張することである。この主張は、毒を飲むことは合理的かという問いに対する直接的な答えにはなっていない。本稿が注目するのは、毒を飲むこと自体の合理性ではなく、毒を飲むことは合理的であるという信念の合理性である。なぜなら、毒パズルにとって重要なのは毒を飲むことを意図できるかどうかであり、そのように意図できるかどうかの決め手は信念にあると考えているからである。このような考えを Holton (2004) が示唆した。この点については 3 節の冒頭で改めて述べる。いずれにしても、本稿の目的は、毒パズルにおける信念の合理性について考察することを通じて、これまで暗黙裡に置かれてきた合理性に関する前提を明らかにし、合理性に対するより広い見方を提案することである。そのために、2 節では、毒パズルの概要を確認し、毒パズルがいくつかの主題から構成されていることを示す。3 節では、毒パズルを構成する主題のなかでも、毒を飲むことは合理的かという主題を中心になされてきた議論を取り上げ、より広い論点のもとで捉え直す。

## 2 毒パズルとは

### 2.1 毒パズルの概要

毒パズルは次のような内容の思考実験である<sup>\*2</sup>。変わり者の大富豪とあなたは取り決めをする。もしあなたが、毒(その毒を飲むと一日苦しむことになるが、毒の効果は致死的でもなければ永久的でもない)を明日の午後に飲むことを、今夜の午前 0 時に意図できたら、それだけで、大富豪は、明日の朝、あなたに 100 万ドルを支払うという取り決めをする。つまり、あなたは、毒を実際に飲まなくても、毒を飲むことが意図

---

<sup>\*1</sup> 私の知る限り、国内には毒パズルを中心的に取り扱っている論文は存在しないが、部分的に取り扱っている論文であれば存在する。それは、柏端 (2008) と杉本 (2015), pp103f である。

<sup>\*2</sup> Kavka (1983)。

できさえすれば、それだけで 100 万ドルが獲得できるのである。

あなたは 100 万ドルを獲得できるだろうか。素朴に考えると、100 万ドルを獲得することなど造作もないことのように感じられる。なぜなら、我々は、普段、様々なことを意図しているので、毒を飲むことを意図するぐらいたやすいように感じられるからである。そして、さらによいことに、あなたは実際に毒を飲まなくても毒を飲むことを意図するだけで 100 万ドルが獲得できることになっている。

さて、あなたが毒を飲むことを今夜の午前 0 時に意図できて、明日の朝 100 万ドルが手に入ったとしよう。100 万ドル獲得後に、あなたは毒を飲むだろうか。おそらく飲まないだろう。なぜなら、毒を飲む時刻には、毒を飲む理由はない(なぜなら、すでに 100 万ドルが手に入っている)し、毒を飲まない理由もある(なぜなら、毒を飲むと一日苦しむことになる)からである。よって、毒を飲む行為は不合理な行為だということになる。不合理だと分かっている行為を本気で意図することなどできるだろうか。

以上のように考えていくと、100 万ドルを獲得することが難しく思えてくる。これが、毒パズルのパズルたるゆえんである。たとえあなたが 100 万ドルを獲得するために一日ぐらい苦しんでも構わないと考えていたとしても、100 万ドルを獲得することはできないだろう。

Kavka (1983) は三つの前提を明示的に置いている。第一に、毒パズルにおいては、毒を飲んで賞金を獲得するために、外部の力を借りることは禁止されている。たとえば、毒を飲まなければ自分のことを殺すよう殺し屋に依頼することは禁止されている<sup>\*3</sup>。第二に、毒パズルにおいては、毒を飲むことに関わる諸条件が十分に信頼できると想定されている。たとえば、大富豪がサインした法的な書類がきちんとしたものであるということは調査済みである<sup>\*4</sup>。第三に、カフカは意図を「行為の理由に基づく行為の傾向性」として捉えている。なぜなら、もし意図が行為の傾向性ではなく心的な行為や制御可能な決定だとしたら、毒を飲むことを今夜の午前 0 時に意図することは容易なので、パズルが成り立たなくなってしまうからである。

## 2.2 毒パズルを構成するいくつかの主題

以上のように毒パズルを概観していくと、100 万ドルを獲得することなど簡単だという先程の素朴な直観が崩れ去り、これは無理な取引だと感じられるようになる。それと同時に、哲学的な興味を持った読者であればもっと大きな当惑を感じているかもしれない。毒パズルでは、そもそも何を哲学的に問題にすべきなのかが判然としない。その原因の一つに毒パズルがいくつかの主題から構成されていることがあるだろう。そこで、ひとまとめに扱われている主題をいくつかに分けることにしよう。

<sup>\*3</sup> この制限によってブロックされているのは次のような事例である。あなたが、毒を飲むことを意図するためには、毒を飲むことが合理的だと思わなければならないと考えたとしよう。このとき問題になるのは、毒を飲むことが不合理だということに気づいていると、毒を飲むことが合理的だと思いが難しくなり、意図できなくなることである。そこで、殺し屋の出番である。殺し屋を雇って、もし自分が毒を飲まなかったら殺してもらうことにすれば、100 万ドルを獲得したあとも、毒を飲む理由がなくなることはない。なぜなら、あなたは殺されたくないと思っているからである。

<sup>\*4</sup> この制限によってブロックされているのは次のような事例である。あなたが「大富豪は毒を飲むことを意図するだけで 100 万ドルをくれると口では言っていたが、本当は、毒を飲まないと 100 万ドルをもらうことはできないのだ」などと思い込むことによって、毒を飲む理由をつくり出す事例である。

私の見立てでは、以下のような主題が毒パズルを構成している\*5。これらの主題が毒パズルの豊かさを示している。

- (1) 毒パズルという思考実験は原理的に可能か。
- (2) 毒を飲むことを意図できるか。
- (3) 毒を飲むことは合理的か。

注意を二つしておこう。第一に、これらの主題が、毒パズルを構成する主題のすべてではない。第二に、これらの主題は互いに独立しているわけではない。たとえば、(2) と (3) は密接に関わり合っている。本稿でも、合理的な行為だけが意図できるのか、あるいは、不合理な行為は意図できないのかということが問題になる。とはいえ、(1) から (3) のようにひとまず区別しておくことは有用である。

上に挙げた三つの主題をそれぞれ簡単に見ておく。

**(1) 毒パズルという思考実験は原理的に可能か**\*6。毒パズルのような奇妙な事態が起きることは原理的にないのだから、そもそも論じるに値しないという反論がありうる。G・カフカが毒パズルを提出した時代には、脳の状態から意図を読み取る機器など空想の産物でしかなかっただろう\*7。しかし、近年、脳の状態を読み取って機械を操作したり、機械から脳に直接刺激を与えることによって情報を伝達したりする Brain-Machine Interface が実証的に研究されている。毒パズルという思考実験が原理的に可能かどうか、脳科学や Brain-Machine Interface などの知見も参照しつつ考えることは、意図に対する哲学の見方と脳科学の見方の内実や、それらの間の関係について考えることにもなる。

**(2) 毒を飲むことを意図できるか**。毒パズルについて論じ始めると、まず問題になるのが、被験者が 100 万ドルを獲得できるかどうかということ、つまり、毒を飲むことを意図できるかどうかということである。素朴に考えれば、100 万ドルを獲得することは簡単であるように感じられる。しかし、100 万ドルを獲得するためには、不合理だと思っている行為を意図する必要があるように思われた。2.1 節では、不合理な行為は意図できないという前提に立って話を進めたが、議論の余地はある。そもそも我々は、「意図」がどのようなものなのかははっきりと理解しているわけではない。毒を飲むことを意図できるかどうか考えることは、結果として、意図がどのようなものなのかを考えることにもなる。

**(3) 毒を飲むことは合理的か**。本稿が最終的に問題にするのは、毒を飲むことは合理的であるという信念の合理性である。ここではひとまず、不合理な行為は意図できないということを前提にしよう。すると、もしあなたが「毒を飲むことが不合理だ」と思っているならば、毒を飲むことが意図できなくなり、100 万ドルも獲得できないことになる。そこで、あなたは「毒を飲むことが合理的だ」と思うことに決め、そう思うことに成功したとしよう。この場合も問題が生じる。あなたは「毒を飲むことが合理的だ」と 100 万ドル獲得後も思い

---

\*5 カフカ自身は (2) を明示的に問題にしている。

\*6 佐藤広大 (2018) で詳しく論じている。

\*7 実際、Gauthier (1998), p.47 では、被験者の脳の状態から、毒を飲む意図を読み取る機器の存在可能性に対して疑義が呈されている。

続けることができるだろうか。そのように思い続けることはできないので、毒を飲むことも意図できないだろう。したがって、本稿でポイントになるのは、被験者が100万ドル獲得後についてどのように考えるのか、あるいは、100万ドル獲得後についてそもそも考えないことは可能かということである。

### 3 毒を飲むことは合理的か

本節では、毒パズルを構成する主題のなかでも、毒を飲むことは合理的かという主題を中心になされてきた議論を取り上げる。その主題に対して、本稿は、毒を飲むことが合理的だという信念を、意図する前に形成することは合理的だと答える。そのために、3.1節では、不合理派(毒パズルにおいて毒を飲むことが不合理だと主張する立場)の議論を確認する。次いで、3.2節では、合理派(毒を飲むことが合理的だと主張する立場)のいくつかの議論を批判的に検討する。3.3節では、本稿の立場を打ち出し、擁護する。

本稿のタイトルにあるとおり、これまで先行研究の多くは、毒を飲むことが合理的かどうかについて争ってきた。それに対して、R・ホルトンは、毒を飲むことの合理性ではなく、毒を飲むという意図を考え直すことの合理性を問題にした。ホルトンの問題設定は示唆的である。なぜなら、たとえ毒を飲むことが不合理だったとしても、被験者が毒を飲むことが合理的だと思うことが可能(あるいは合理的)であれば、毒を飲むことを意図できるからである。そして、もし毒を飲むことを意図できれば、100万ドルを獲得するという最善で十分な理由がある選択肢をとったことになる。ホルトンの示唆のもとで先行研究を捉え直すと、実のところそのほとんどが、被験者が意図を形成する前に毒を飲むことについて十分熟慮することを前提にしたうえで、毒を飲むことが合理的だと被験者が思うかどうかを争っていたことになる<sup>\*8</sup>。そして、私の見立てでは、先行研究の論争は、もし被験者が意図形成前に十分に熟慮すれば、毒を飲むことを不合理だと思わざるをえず、意図を形成できないということを示唆している。しかし、そもそも被験者は十分熟慮しなければならないのだろうか。すなわち、被験者が十分熟慮するということは毒パズルにおいて本当に合理的なのだろうか。以下で、私は、毒を飲むことが合理的だと思うことができる程度のほどよい熟慮こそがむしろ合理的であり、そのような熟慮を行う被験者であれば毒を飲むことを意図できると主張する。

#### 3.1 不合理派の議論

不合理派は、どのようなことを根拠に、毒を飲むことが不合理だと主張するのだろうか。

たとえば、毒パズルを提出したカフカ自身は、毒を飲むことが不合理である<sup>\*9</sup>理由を次のように説明している。毒を飲む予定の時刻(明日の午後)には、大富豪から銀行口座にすでに100万ドルが振り込まれているかいないかのどちらかなので、あなたが実際に毒を飲むかどうかは100万ドルが振り込まれているか

<sup>\*8</sup> 「毒を飲むことは合理的か」という問いに取り組むことが、結果的には「毒を飲むことは合理的だと信じることは主体にとって合理的か」という問いに取り組むことになるのはなぜだろうか。「命題 $\phi$ が真ならば、 $\phi$ と信じることは主体にとって合理的である」ということが前提にされているからかもしれない。この前提を認め、かつ、毒を飲むことが合理的だという命題が真ならば、毒を飲むことは合理的だと信じることは主体にとって合理的であるということが帰結する。

<sup>\*9</sup> 正確には Kavka (1983), p.35 では「不合理」という語は用いられておらず、代わりに「行為する理由が全くない(you have no reason to act)」という語が用いられている。

どうかに影響を与えることはない。したがって、毒を飲む理由はないので、毒を飲むことは不合理である (Kavka 1983, p.34)。

ここまでの議論だけを使って、カフカが主張できるのは、毒を飲むことが不合理だと今 (毒を飲むことを意図しようと様々なことを考慮する時点) から思うことが合理的だということではなく、毒を飲むことが不合理だということまでである。カフカと対立する合理派の人々は、カフカの議論を認めて、毒を飲むことが不合理だということに納得しつつも、一方で、様々なこと (たとえば、毒を飲むことが不合理だと思っていると、毒を飲むことが意図できなくなってしまうように感じられるということなど) を考慮すると、100 万ドルを獲得するためには、毒を飲むことが合理的だと思わざるを得ないので、毒を飲むことは合理的だと今思うことは合理的だと主張する余地がある。すなわち、問題にすべきは、毒を飲むことが不合理だと今思うことが合理的かどうかということである。

このことを考えるために、カフカの議論で何が暗に前提にされているかを確認しておきたい。カフカが答えを与える際に鍵となっているのが「反映原理<sup>\*10</sup>」と呼ばれる規範的な指針である。反映原理は、以下のような帰結を持ち、それがカフカの議論で暗に前提にされている第一のものでもある。

【CRP】ある命題  $\varphi$  を未来の時点  $t$  で確信することが今から分かっているのだとすれば、 $\varphi$  を今から確信するべきだ。

この帰結を用いて、カフカは、毒を飲むことは合理的だと今思うことは不合理だと答えるだろう。なぜなら、今持っている証拠 (大富豪と取り交わした契約の内容など) から、毒を飲むことが不合理だと明日の午後に確信することがすでに分かっているからである。このとき CRP を用いる<sup>\*11</sup>と、「毒を飲むことが不合理だ」と明日の午後に確信することが今から分かっているのだから、「毒を飲むことが不合理だ」と今から確信するべきだということになる。したがって、たとえ合理派が、「毒を飲むことが合理的だ」と今思うことは合理的だと反論したとしても、カフカに言わせれば、「毒を飲むことが不合理だ」と明日の午後に確信することが今から分かっているのだから、「毒を飲むことが不合理だ」と今から思うべきなのである。

第二に、カフカの議論では、被験者が毒パズルのルールを理解しているということが前提にされている。もし被験者が毒パズルのルールを理解していなければ、毒パズルという思考実験がそもそも成立しなくなってしまう。たとえば、被験者が大富豪の言葉を十分に理解できず、100 万ドルを獲得するための条件が分かっていなければ、毒パズルは成立しない。よって、カフカが、毒パズルを成立させるために、被験者

<sup>\*10</sup> van Fraassen (1984), p.244 では、次のような反映原理が提出されている。 $P_{a,t}(\varphi | P_{a,t+t'}(\varphi) = r) = r$ 。「 $P_{a,t}(\varphi) = r$ 」は、主体  $a$  が命題  $\varphi$  に対して  $r$  の信念度合を時点  $t$  に持つことを表わす。反映原理を日常言語で表わすと「 $\varphi$  に対する自分の信念度合が今から  $t'$  後に  $r$  であるということが分かっているのだとすれば、 $\varphi$  に対する現在の信念度合もまた  $r$  であるべきだ」ようになるだろう。反映原理は、通時的に合理的であるための条件であり、従わなければ、賭けにおいてダッチブックを作られて損することになる。本文中の「CRP」は、 $P_{a,t}(\varphi | P_{a,t+t'}(\varphi) = 1) = 1$  と表現でき、 $r=1$  の場合の反映原理である。

<sup>\*11</sup> 反映原理をめぐる多くの議論がなされている。たとえば、反例として Elga (2000) などが知られている。議論の多さが示すように、反映原理は無自覚に前提にされてよいものではない。

が毒パズルのルールを理解しているという前提を暗に置いていたことは正当化される。

第三に、カフカの議論では、被験者が毒パズルにおいてどのような場合が生じうるかをあらかじめ計算するということが前提にされている。カフカの巧妙な議論の進め方のために、我々はこの前提を無自覚のうちに受け入れてしまっている。私は、この前提をある仕方で受け入れることには問題があると考えている。Kavka (1983) では、毒パズルがいかによりパズルたりえているかを描くために、被験者の視点が次のように利用されている。被験者は、大富豪から、毒を飲むことを意図するだけで 100 万ドルが獲得できるという取引を持ちかけられ、簡単に 100 万ドルが獲得できそうだと最初は喜ぶが、等々。被験者の視点から毒パズルをこのように説明する都合上、毒パズルにおいてどのような場合が生じうるかを被験者があらかじめ計算するとカフカは想定している。つまり、被験者が、不合理な行為は意図できないという前提に基づいて、毒を実際に飲む必要があるかを、時系列を追って検討していき、毒パズルにおいてどのような場合が生じうるかを計算すると想定されている。

カフカのこの想定を、我々はどのように解釈できるだろうか。少なくとも三つの解釈が可能だろう。

第一に、毒パズルがいかによりパズルたりえているのかを説明するために、カフカはそうのように想定していたと解釈できるだろう。カフカは、100 万ドルを獲得できるかどうか被験者が思いをめぐらせているさまを描写することを通じて、毒パズルがいかによりパズルたりえているかを示した。毒パズルを説明するためや、そもそも毒パズルを成り立たせるために、被験者が毒パズルにおいてどのような場合が生じうるかを計算すると想定する必要がある、そのためだけに、カフカが、このように想定しているのであれば何の問題もない。つまり、第一の解釈をとった場合、カフカは、毒パズルをパズルとして成り立たせるためだけに、このような被験者を想定したのであって、我々が被験者のように振る舞うべきだとまでは主張していないことになる。

第二に、もし我々が毒パズルの被験者になったら、どのような場合が生じうるかを「通常」は計算するとカフカが考えていたために、被験者が毒パズルにおいてどのような場合が生じうるかを計算すると想定していたと解釈できるだろう。我々が毒パズルに触れたとき、毒を飲むことは不合理だというカフカの主張がもっともらしく思える要因の一つとして、もし自分たちが毒パズルの被験者になったら、どのような場合が生じうるかを計算してしまうだろうと予想できることが挙げられる。毒パズルのようなゲーム形式(被験者の選択肢があらかじめ決まっていて、それぞれの選択肢に応じた賞罰も決まっている形式)の思考実験では、被験者が可能な全通りの選択肢とそれぞれの選択肢に応じた賞罰を計算していることが多くの場合前提にされている。そのため、カフカが、毒パズルにおいて被験者はどのような場合が生じうるかを通常は計算すると想定していたとしても問題がないように感じられるかもしれない。しかし、このときの「通常」という語は、どのような意味なのだろうか。たしかに、ゲーム形式の思考実験に直面すると、どのような場合が生じうるかを計算したくなってしまうが、そのような状態は「通常」なのだろうか。

このような疑問に対して、「被験者がどのような場合が生じうるかを通常は計算する」と言うときの「通常」という語は「被験者が合理的である場合」という意味だという答えがありうるだろう。つまり、第三の解釈として、合理的な被験者であればどのような場合が生じうるかを計算するとカフカが想定していたと考えることができる。しかし、この「合理的な」という語が曲者である。どれくらいのことを計算していれば被験者が合理的



だと言えるのだろうか.

結論を先取りすれば、私が示したい見方とは、状況に応じてほどよい考え具合、あるいは、計算具合があるというものである. この見方は、考えれば考えるほど合理的になるという一般に抱かれがちな見方とは異なるものである. 本稿では、第三の解釈 (合理的な被験者であれば、毒パズルにおいてどのような場合が生じうるかを計算するとカフカは想定したという解釈) に絞って検討していく. なぜなら、もし第一の解釈通りであれば特に問題はなく、第二の解釈通りであれば「通常」という語が漠然としすぎていて議論にならないからである.

## 3.2 合理派の議論

本節では、合理派の議論を批判的に検討していく. 拒否している不合理派の前提に応じて、合理派をいくつかの立場に分けることができる.

### 3.2.1 意図の本性から、毒を飲むことは合理的だと主張する立場

不合理派が前提にしている意図理論を拒否して、独自の意図理論をもとに、意図の本性から毒を飲むことは合理的だと主張する立場がある. 様々なものが意図の本性の候補となりうるが、意図の本性は自分との約束であることだと主張する立場 (「約束説」) を取り上げる<sup>\*12</sup>.

約束説を K・レヴィが紹介している (Levy 2009, pp.273f). レヴィによれば、その立場は、毒を飲むという合理的な<sup>\*13</sup>意図を形成したとき、毒を飲む約束を自分とすることになるため、毒を飲むことは、すなわち、自分との約束を果たすことは、合理的だと主張する.

レヴィ自身は、約束説を論敵としていて<sup>\*14</sup>、まず「前向きの理由 (forward-looking reasons)」と「後ろ向きの理由 (backward-looking reasons)」の区別を導入する. ある行為 A をするための前向きの理由とは、A がもたらすよい帰結に基づいて、A を支持するような考慮事項のことである. 一方、A をするための後ろ向きの理由とは、A よりも以前に起こったことに基づいて、A を支持するような考慮事項のことである. 「A よりも以前に起こったこと」とは、たとえば、約束をしたことや、悪い行いをしたことなどである. この区別を踏まえて、レヴィは約束説に反論する. 約束説によれば、毒を飲むという意図を形成するとき、毒を飲むという約束を自分としたことになる. この約束を、レヴィは、道徳的なものではなく、功利的なものだと考える. なぜなら、毒を飲むという約束は、他人との間でなされてはいない<sup>\*15</sup>し、お金を得るという利己的な目的からなされてもいるからである. そして、レヴィは、もしこの約束が功利的なものであれば、痛みを最小化するた

<sup>\*12</sup> カフカ自身、毒を飲む約束を自分とすることに言及し、被験者はそのような約束を守ることができない人物だと想定している (Kafka 1983, pp.34f).

<sup>\*13</sup> 毒を飲むことを意図すれば 100 万ドルが獲得できるので、毒を飲むことを意図するのは合理的である.

<sup>\*14</sup> Levy (2009), pp.270f, 273f.

<sup>\*15</sup> 自分との約束を他人に向けて意図的に宣言すれば、そのときには、約束通りに行為すべきある種の義務が発生するかもしれない. 毒パズルにおいて、被験者に毒を飲む意図があるかどうかは、大富豪が脳走査装置で確かめることになっている. そして、被験者もそのことを知っている. だとすると、毒を飲む義務が被験者にはあるのではないかという疑問が浮かぶかもしれない. この疑問に対して、私は、そのような義務はないと答えたい. なぜなら、意図が読み取られるという出来事を、被験者は承認しているだけで、自分から進んで行っていないからである.

めの前向き理由に上回られてしまうと主張する。なぜなら、毒を飲むという約束は功利的なものであるにもかかわらず、前向き理由ではなく、後ろ向き理由だからである。つまり、毒を飲むという約束の根底には、毒を飲むことによって長期的な利益を得たいという前向きな動機が存在している。にもかかわらず、毒を飲むという約束自体は、長期的な利益を得るためになされた後ろ向き理由である。ここでのポイントは、毒を飲むという約束の根底にある動機（毒を飲むことによって長期的な利益を得たいという前向きな動機）は毒を飲む前に満たされてしまうので、痛みを最小化するための前向き理由が毒を飲むという約束を守るための後ろ向き理由を上回ることである。以上から、毒を飲まないことが考慮事項によって最終的に支持されるとレヴィは主張する。

### 3.2.2 計画全体が合理的だから、その計画の一部である毒を飲むことも合理的だと主張する立場

不合理派が前提にしていなかった「計画」という概念を導入することによって、計画全体が合理的だからその一部である毒を飲む行為も合理的だと主張する立場の議論がある。次に、この立場を見ていく。

D・ゴティエはおおよそ次のように論じている<sup>\*16</sup>。毒パズルにおいて可能な計画は、毒を飲むことを意図し毒を飲むという計画（「計画 1」）と、毒を飲むことを意図せず毒を飲まないという計画（「計画 2」）の二つしかない。このとき、毒を飲むことを意図し毒を飲まないという計画（「計画 3」）が可能であり<sup>\*17</sup>、むしろこの計画こそ最善なのではないかという疑問が生じるかもしれない。しかし、計画 3 は、自分がやらないだろうと思っている行為を意図することはできないという意図の本性から不可能である。よって、可能な計画は、計画 1 と計画 2 だということになる。計画 1 をとれば一日苦しむ代わりに 100 万ドルが手に入るのに対して、計画 2 をとれば苦しむことはないが 100 万ドルが手に入ることもない。二つの計画の快苦の総和を比べると、計画 1 の方がより望ましいので合理的であり、合理的な計画 1 の部分である毒を飲む行為も合理的である。以上のようにゴティエは主張している。

計画全体が合理的だからその一部である毒を飲む行為も合理的だと主張するこの立場には説得的な部分がある<sup>\*18</sup>。たしかに、もし計画全体が合理的であれば、その一部も合理的になるだろう。たとえば、次のような事例（「ぬりつぶし事例」）を考えてみよう。子どもが白い画用紙を赤一色で塗りつぶしている。画用紙を赤一色でぬりつぶすというこの行為だけを見ていると、この行為が不合理なもののように感じられるかもしれない。ところが、その子どもは、実は、赤くぬりつぶした画用紙を組み合わせで大きなりんごの絵をつくるという計画を実行している最中だったのだ。大きなりんごの絵をつくるという計画全体は合理的だとして。たとえば、その子どもは、りんご好きの自分の母親のために大きなりんごの絵をつくる方法として、このように白い画用紙を赤一色でぬりつぶしていくのが最善だと考えていた。このとき、大きなりんごの絵をつくるという計画全体が合理的なので、その計画の一部である、白い画用紙を赤一色でぬりつぶすとい

<sup>\*16</sup> Gauthier (1994), p.708; (1998), pp.47f.

<sup>\*17</sup> 可能な計画の候補が、計画 3 の他にもう一つ存在する。その計画（「計画 4」）とは、毒を飲むことを意図しないで毒を飲むという計画である。意図しないのに行為するという、おそらく習慣的な行為の可能性は興味深い。しかし、本文中では、意図の本性について詳細に考察する必要があるため、計画 4 を取り上げなかった。ゴティエ自身も計画 4 については考察していない。

<sup>\*18</sup> 以下の論点を金杉武司氏に指摘していただいた。

う行為も合理的だろう。なぜなら、白い画用紙を赤一色でぬりつぶすという行為には、それを実行する理由（白い画用紙を赤一色でぬりつぶせば、大きなりんごの絵をつくるという合理的な計画が成り立つという理由）があるし、やめる理由もない（たとえば、ぬりつぶすのをやめるように母親に注意されてもいない）からである。

しかし、毒パズルは〈ぬりつぶし事例〉のようにはいかない。毒パズルでは、計画 1（毒を飲むことを意図し毒を飲むという計画）はそもそも計画として成り立たない。その理由は、計画における目的の位置と関係している。先程の〈ぬりつぶし事例〉では、子どもの計画は、大きなりんごの絵を完成させるという目的を終点としていた。一方、計画 1 は 100 万ドルの獲得を目的としているが、その目的は計画の終点ではなく途中にある。計画 1 のように、その途中に目的がある計画は、計画として成り立たないと思われる。なぜなら、計画の途中にある目的が達成された時点で、その計画は終わってしまうように感じられるからである。我々は、通常、まだ達成されていない未来の目的のために努力するのであって、すでに達成されてしまった過去の目的のために努力することはない<sup>\*19</sup>。目的達成のために行っていた過去の努力を、習慣や惰性から、目的達成後も続けることはあるが、そのような努力の継続は今問題になっている「過去の目的のための努力」ではない。したがって、ひとたび計画の途中で目的が達成されてしまえば、その目的のために努力することはないだろう。つまり、計画を進めていくための動機（目的を達成したいという動機）を失うことになるだろう。そして、計画 1 のような、目的が途中にある計画に関して、計画を進めていくための動機を途中で失ってしまうことがあらかじめ分かっていたとしたら、そもそも計画 1 を立てることができなくなり、合理的な計画 1 の部分である毒を飲む行為は合理的だという主張も成り立たなくなるはずである。

### 3.2.3 毒を飲む意図を考え直さない傾向性が合理的だから、毒を飲むことも合理的だと主張する立場

以下では、まずホルトンの議論を取り上げる。ただし、ホルトン自身は、毒を飲むことが合理的だとまでは主張しない。彼が主張するのは、毒を飲む意図を考え直さないことが合理的だということである。しかし、合理派はホルトンの立場を援用して、毒を飲む意図を考え直さない傾向性が合理的だから毒を飲む行為も合理的だと主張できる（「傾向性説」）。この立場は、被験者は毒を飲む意図を 100 万ドル獲得後には放棄するはずだという不合理派の前提を拒否する。私は、この傾向性説が、修正する必要はあるが、合理派の既存の立場のなかで最も重要な示唆を与えていると考える。

ホルトンは、どのように主張しているのだろうか。ホルトンによれば、毒を飲む意図を 100 万ドル獲得後に考え直さないことが合理的なのは、考え直さなければ 100 万ドルを獲得できるからである。そして、そのようにするためには、あるタイプの意図を考え直さない傾向性を持っている必要があるとホルトンは主張する。「あるタイプの意図」とは、それを考え直さなければ、利益が生まれるようになるところの意図である。毒パズルの場合には、毒を飲む意図である（Holton 2004, pp.528f）。

合理派は、ホルトンのこの議論を援用して次のように傾向性説を主張できる。第一に、毒を飲む意図は

<sup>\*19</sup> 同様の指摘を Bratman (1998), p.66 も行っている。M・ブラットマンによれば、ゴティエは「行為者の因果的な制御下にはもはやない要素を含むという特徴を持った行為のコースを扱っている。このような行為のコースを扱っているということは、我々の行為者性にとって時間的な位置や因果的な位置が重要であることを正当に評価していないということだと思われる」。

合理的である。なぜなら、毒を飲むことを意図すれば、100 万ドルが獲得できるからである。第二に、毒を飲む意図を 100 万ドル獲得後に考え直さない傾向性は合理的である。なぜなら、ホルトンも主張しているように、毒を飲む意図を 100 万ドル獲得後に考え直さない傾向性を持っていれば、100 万ドルが獲得できるからである。よって、毒を飲む合理的な意図と、毒を飲む意図を考え直さない合理的な傾向性から生じる、毒を飲む行為も合理的である。

傾向性説のこの議論に対して、私は次のような指摘をしたい。たしかに、毒を飲む意図や、毒を飲む意図を 100 万ドル獲得後に考え直さない傾向性は、合理的なのかもしれない。しかし、意図と傾向性が合理的だからといって、毒を飲む行為までもが合理的だということには必ずしもならない。

なぜだろうか。毒パズルにおいて、毒を飲む意図と、毒を飲む意図を考え直さない傾向性が合理的だと仮定しよう。このように仮定しても、もし被験者が毒を飲む行為が不合理だという信念を 100 万ドル獲得前に持っていれば、毒を飲む意図(そして毒を飲む行為)を 100 万ドル獲得後に考え直さずに、被験者は毒を飲む行為が不合理だと思うことがあるだろう。たとえば、被験者は毒を飲む行為が不合理だということを思い出して、毒を飲まないかもしれない。もし被験者が毒を飲むことを意図する前からこのことに気づいていたら、毒を飲むことを意図できないだろう。したがって、たとえ、毒を飲む意図や、毒を飲む意図を考え直さない傾向性が合理的だったとしても、毒を飲む行為までもが合理的だということには必ずしもならない。

### 3.3 本稿の立場

3.2.3 で見た傾向性説の力点は「意図」と「意図を考え直さない傾向性」にあった。それに対して、私は、毒パズルにおいて毒を飲むことが合理的かどうか論じる際に鍵となるのは「意図」ではなく「信念」だと主張する<sup>\*20</sup>。そして、毒を飲むことが合理的だという信念を、毒を飲むことを意図する前に形成することは合理的だと主張する立場を展開していく。これが本稿の支持する立場である。この立場は、合理的な被験者は毒パズルにおいてどのような場合が生じうるかをすべて計算するという前提を拒否する。

傾向性説に従って、毒を飲む意図を考え直さなかったとしよう。それでも、もし被験者が毒を飲む行為が不合理だという信念を 100 万ドル獲得前に持っていれば、毒を飲む意図を考え直さずに、毒を飲む行為が不合理だということを 100 万ドル獲得後に思い出すかもしれないので、毒を飲むことを意図できないのだった。であれば、毒パズルにおいて毒を飲むことが合理的かどうか論じる際に鍵となるのは「意図」ではなく、どのように考えるか、すなわち「信念」だということになるのではないだろうか。

私は、ある信念を支える証拠自体を制御できるか、あるいは、どの証拠を探すか決めることができるなら、ある信念を抱くかどうかを間接的に制御できると考えている<sup>\*21</sup>。このような考えを、毒パズルに適用すると、

<sup>\*20</sup> ホルトン自身に対しては、毒パズルを論じる際に鍵となるのは「意図の持続」よりも「意図の形成」であると反論していることになる。

<sup>\*21</sup> この考え方は、合理性をめぐる哲学の議論においては伝統的なものである。たとえば Feldman (2001) を参照されたい。本文中の「ある信念を支える証拠自体を制御できる[...]なら、ある信念を抱くかどうかを間接的に制御できる」ことの具体例として R・フェルドマンは次のようなものを挙げている (Feldman 2001, pp.81f)。あなたはスイッチを操作することによって電気を点けたり消したりできるとしよう。すると、あなたは電気を点けることによって、電気が点いているという

どのように主張することになるだろうか。熟慮するという心的行為は、意志によって、途中で止めるなど制御できるので、熟慮するという行為から生じる信念の形成も間接的に制御できるということになる。よって、毒パズルにおいて、100 万ドルを獲得するためには毒を実際に飲む必要がありそうだという結論に、つまり、毒を飲むことが合理的だという結論にたどり着いたら<sup>\*22</sup>、熟慮するのを意図的に止めてしまえばよいのである<sup>\*23</sup>。そのように止めることができれば、100 万ドルを獲得することができる。そこからさらに考えを進めて、毒を 100 万ドル獲得後に飲むことが合理的かどうか熟慮してはいけない。なぜなら、毒を飲むことが不合理だという結論にひとたびたどり着いてしまうと、そのような信念を毒パズルにおいて改訂することは困難だからである。

よって、私の立場は次のようなものになる。100 万ドル獲得のためには毒を飲むことが必要だという結論にたどり着いたら、それ以上考えるのを意図的に止めてしまえば、毒を飲むことが合理的だという信念を意志によって間接的に形成でき 100 万ドルも獲得できる。したがって、毒を飲むことが合理的だという信念を、意図する前に形成することは合理的である。たとえ、被験者がその信念を形成することが最善である理由を説明できなかったとしても、その信念を形成することは合理的である。

私の立場にとって鍵となるのは、毒を飲むことが不合理だという信念を、毒を飲むことを意図する前に形成することが不合理だという主張である。その一方で、先行研究は、次のことを前提にしてきた。被験者が、大富豪から取引を持ちかけられたときに、思考をめぐらせて、毒を飲むことが不合理だという信念を形成するのは合理的だということを前提にしてきた。しかし、私の見立てでは、そのような被験者は考えすぎている。100 万ドルを獲得するために毒を飲む必要がありそうだと思うことができたならば、それ以上考える

信念を抱くことができる。

<sup>\*22</sup> 毒を飲むための理由がないことと、毒を飲まないための理由があることが被験者にとってあまりにも明白なので、毒を飲むことが合理的だという結論にたどり着くことが想像しにくいと反論されるかもしれない。この反論に対して、私は次のように答えたい。毒を飲むための理由がないことと、毒を飲まないための理由があることが被験者にとって明白であるように感じられるのは、カフカの巧妙な議論の進め方によって、どのような場合が生じうるかを被験者はあらかじめ計算するという前提を暗に認めさせられているからだ。3.1 節でも述べたように、カフカは、毒パズルがいかにパズル तरीえているかを示すために、100 万ドルを獲得できるかどうか被験者が思いをめぐらせるさまを描写している。この描写が巧妙なのは、この描写を通じて読者が毒パズルについて理解するので、カフカが描いたように思考を巡らせることが当然であり、毒を飲むための理由がないことなどが明白だという印象を受けてしまうからである。しかし、毒を飲むための理由がないことなどを理解するためには熟慮する必要があるというのが私の主張である。

この主張は、次のようにさらに再反論されるかもしれない。毒を飲むための理由がないことなどを理解するために熟慮する必要があったとしても、そのような熟慮はほんの少しで十分なので、毒を飲むための理由がないことは依然として明白だと。この再反論に対して、私は次のように答えたい。3.1 節でも述べたように、毒パズルのようなゲーム形式の思考実験に直面すると、どのような場合が生じうるかをつい計算したくなってしまうので、毒を飲むための理由がないことは明白だと感じられるのかもしれない。しかし、そのように計算することは毒パズルにおいては不合理である。ただし、毒パズルにおいて計算しないようにするためには、かなり奇妙な仕方であるかもしれない。被験者が毒パズルのルールを理解している(被験者が大富豪の言葉を十分に理解している等)というカフカの前提を受け入れたうえで、計算しないようにすることが可能であれば(ここでは具体的なメカニズムまで明らかにすることはできないが)、本文中での私の議論は成立する。

<sup>\*23</sup> 熟慮を意図的に止めることを可能にする具体的なメカニズムの解明は重要な課題だが、ここでは取り組むことができない。しかし、次のようなものがそうしたメカニズムの候補となりうるだろう。第一の候補は、ややこしくなったら、あるいは、パラドキシカルな感じがしたら細かく考えるのを止めてしまうという実践訓である。第二の候補は、100 万ドルが獲得できなくなることに気づいていたことを忘れさせたり、熟慮させないようにしたりする意図的な抑制である。

必要はないのである。なぜなら、それ以上考えて、毒を飲むことが不合理だと思ってしまったら、そのような信念を改訂することは毒パズルでは困難だからである。

ここで、私は次のように反論されるかもしれない。すなわち、私が想定している被験者は、毒を飲むことが不合理だということにも気づけない不合理な行為者ではないのかと。たしかに、毒を飲むことが不合理だということに気づけない行為者のなかには、不合理な行為者も含まれているだろう。そのような不合理な行為者とは、たとえば、どのようなときでもあまり考えず、一度抱いた信念を変更しないような行為者である。そのような行為者は、毒パズルにおいて 100 万ドルが獲得できたとしても、別の状況では失敗する可能性が高い。一方、私が想定している被験者は、そのような不合理な行為者ではないにもかかわらず、毒パズルではあまり考えないことによって、毒を飲むことが不合理だとは思わないような行為者なのである。すなわち、私が想定している被験者は、状況に応じて、ほどよく考える行為者なのである。もし私が想定しているこの被験者が依然として愚かであるように感じられるのだとしたら、その原因は「考えれば考えるほど人は合理的になる<sup>\*24\*25</sup>」というドグマを受け入れてしまっていることなのかもしれない。人は考えれば考えるほど必ず合理的になるわけではない。ある特定の状況においては、たとえば、毒パズルのような状況においては、考えを進めていくよりも止めてしまう方が合理的だということが成り立ちうるのである。

## 4 まとめ

毒を飲むことは合理的かという問いに対する本稿の答えは次のようなものになる。毒を飲むことが合理的だという信念を、毒を飲むことを意図する前に形成することは合理的だというものである。本稿は、毒を飲む行為が合理的かどうかということを直接問題にするのではなく、むしろ、毒を飲むことが合理的だという信念がどのように形成されるのかということや、そのような信念を形成することが合理的なのかということ

---

<sup>\*24</sup> 「考えれば考えるほどよりよくなる、あるいは、分析哲学者たちはそのように信じがちである。そのような合理主義のせいで、我々はしばしば激しく非難されている。だから、我々は、その主張より弱い次のような主張へと撤退するかもしれない。行為者は考えれば考えるほどより合理的になるということぐらいは少なくとも真ではないのかという主張へと。この論文の主張は、この弱い方の主張ですら強すぎるというものである。考えれば考えるほど行為者がより愚かになっていくような状況が、つまり、あまり考えないことが合理的なコースであるような状況が存在する。」(Holton 2004, p.507).

<sup>\*25</sup> 毒を飲むことは合理的だと思うことは合理的だという私の主張の二つ目の「合理的」という語と、考えれば考えるほど人は合理的になるというドグマの「合理的」という語は異なる意味なので対立しないと反論されるかもしれない。すなわち、Parfit (2001) による対象付与と状態付与という区別を踏まえて以下のように反論されるかもしれない(この区別に対しては Rabinowicz & Rønnow-Rasmussen (2004) が反論している)。私の主張の二つ目の「合理的」という語は、対象付与的合理性(信じられている内容の合理性)ではなく、状態付与的合理性(信じている状態の合理性)を意味している。なぜなら、もしこの「合理的」という語が対象付与的合理性を意味するならば、毒を飲むことは合理的だと思うことは偽な命題を信じているので不合理だということになるからである。一方で、ドグマで用いられている「合理的」という語は対象付与的合理性を意味している。なぜなら、もしこの「合理的」という語が状態付与的合理性を意味するならば、等しく真理に近づくことができるかぎりにおいて考える時間は短ければ短いほうがよいので、ドグマは偽だということになるからだと反論されるかもしれない。

この反論に対して、私は、私の主張とドグマは対立していると応答したい。私が主張しているのは、毒パズルのような事例においては考えることよりも 100 万ドル獲得が優先されるということである。一方、考えれば考えるほど人は合理的になるというドグマが主張しているのは、100 万ドル獲得よりも考えることが優先されるということであるように思われる。

問題にした。

本稿の流れを振り返ると、2.1 節では、毒パズルの概要を確認し、2.2 節では、毒パズルがいくつかの主題(毒パズルという思考実験は原理的に可能か、毒を飲むことを意図できるか、毒を飲むことは合理的か)から構成されていることと、それらの主題がそれぞれどのようなものなのかを確認した。3.1 節では、不合理派が、どのようなことを暗に前提としながら、毒を飲むことが不合理だと主張しているのかを確認した。取り上げた不合理派の前提は次の三つである。第一の前提は「反映原理の帰結 CRP」で、第二の前提は「被験者が毒パズルのルールを分かっているということ」で、第三の前提は「被験者が毒パズルにおいてどのような場合が生じるかを計算しているということ」だった。3.2 節では、合理派のいくつかの立場を批判的に検討した。3.2.1 では、不合理派の意図理論を拒否して、意図は自分との約束であるという独自の理論をもとに、毒を飲むことは合理的だと主張する立場を確認した。この立場に対して、レヴィは、前向きの理由と後ろの向きの理由を区別し、毒を飲むという約束を分析することによって反論した。3.2.2 では、不合理派が前提にしていなかった「計画」という概念を導入することによって、合理的な計画の一部である毒を飲む行為は合理的だと主張する立場を確認した。この立場に対して、私は、この立場が想定している合理的な計画は計画として成り立たないと反論した。3.2.3 では、被験者は毒を飲むという意図を、100 万ドル獲得後には考え直してしまうという不合理派の前提を拒否して、毒を飲む意図を考え直さない合理的な傾向性を被験者が有していれば、毒を飲むことは合理的だと主張する立場を確認した。この立場に対して、私は次のように反論した。すなわち、たとえ毒を飲む意図が合理的で、かつ、毒を飲む意図を考え直さない傾向性が合理的だったとしても、毒を飲む意図を考え直さずに、毒を飲むことが不合理だという信念を 100 万ドル獲得後に持つことがありうると反論した。そこで、3.3 節では、毒を飲むことが合理的かどうか論じる際にポイントとなるのは「意図」ではなく「信念」であり、信念を意志によって形成することが間接的に可能だと主張した。そして、毒を飲むことが合理的だという信念を、意図する前に形成することは合理的だと結論した。加えて、考えれば考えるほど合理的になるというドグマに対して、状況に応じてほどよく考えることが合理的であるという見方を提示した。

## 付記

本稿の内容は、応用哲学会第 9 回年次研究大会(2017 年)での発表に基づいている。

## 謝辞

佐藤達哉氏、徳永和朗氏、新川拓哉氏、八木沢敬氏、アブリオリ研究会の方々、柏端ゼミの方々、応用哲学会年次研究大会発表会場の方々、査読者の方々から有益なコメントを頂いた。心よりお礼申し上げます。

## 参考文献

- [1] Ainslie, G. 2001. *Breakdown of Will*. Cambridge University Press. [邦訳:『誘惑される意志—一人はなぜ自滅的行動をするのか』山形訳, NTT 出版, 2006. ]
- [2] Bratman, M. 1998. Toxin, temptation, and stability of intention. Coleman, J. L. & Morris, C. W., eds. *Rational Commitment and Social Justice: Essays for Gregory Kavka*: 59-83. Cambridge University Press.
- [3] Chignell, A. 2017. The ethics of belief. *The Stanford Encyclopedia of Philosophy*.
- [4] Elga, A. 2000. Self-locating belief and the sleeping-beauty problem. *Analysis*, 60: 143-147.
- [5] Feldman, R. 2001. Voluntary belief and epistemic evaluation. Steup, M. ed. *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*: 77-92. Oxford University Press.
- [6] Gauthier, D. 1994. Assure and threaten. *Ethics*, 104: 690-721.
- [7] ———. 1998. Rethinking the toxin puzzle. Coleman, J. L. & Morris, C. W., eds. *Rational Commitment and Social Justice: Essays for Gregory Kavka*: 47-58. Cambridge University Press.
- [8] Holton, R. 2004. Rational resolve. *The Philosophical Review*, 113: 507-535.
- [9] Kavka, G. 1983. The toxin puzzle. *Analysis*, 43: 33-36.
- [10] Levy, K. 2009. On the rationalist solution to Gregory Kavka's toxin puzzle. *Pacific Philosophical Quarterly*, 90: 267-289.
- [11] Parfit, D. 2001. Rationality and reasons. Egonsson, D. Josefsson, J. Petersson, B. & Rønnow-Rasmussen, T., eds. *Exploring Practical Philosophy: From Action to Values*: 17-39. Ashgate.
- [12] Rabinowicz, W. & Rønnow-Rasmussen, T. 2004. The strike of the demon: on fitting pro-attitudes and value. *Ethics* 114: 391-423.
- [13] van Fraassen, B. 1984. Belief and the will. *Journal of Philosophy*, 81: 235-256.
- [14] 柏端達也. 2008. 「アクラシア, 高階性, 通時的合理性」. 『科学哲学』 41: 45-58.
- [15] ———. 2014. 「欲求の反映と人称的バイアス」. 『哲学』 133: 1-21.
- [16] 佐藤広大. 2018. 「毒パズルと脳走査装置」. 『新進研究者 Research Notes』 1: 26-34.
- [17] 杉本俊介. 2015. 「'Why Be Moral?'問題の再検討」. 京都大学. 博士論文.

## 著者情報

佐藤広大(慶應義塾大学)